LB 1590
.T6
Copy 1

Class LB1590

Book T6

Copyright No.

COPYRIGHT DEPOSIT.

HANDWRITING

463
965

By EDWARD L. THORNDIKE

Professor of Educational Psychology, Teachers College
Columbia University

Reprinted from Teachers College Record, Vol. 11, No. 2, March 1910.

PUBLISHED BY
Teachers College, Columbia University
NEW YORK CITY
1912

LB1590
.16

Copyright 1910
Copyright 1912

© Cl. A 314733
No 1

CONTENTS

SECTION		PAGE
1.	Introduction	1
	PART I. THE MEASUREMENT OF THE QUALITY OF HANDWRITING	
2.	The Construction of a Scale for Quality of Handwritings by Children in Grades 5 to 8	4
3.	The Nature of the Scale, including Qualities or Degrees of Merit from that of Copy-Book Models down	7
4.	Criticisms of the Scale	8
5.	The Uses of the Scale	17
6.	A Scale for Quality in Adult Women's Writing	19
7.	The Derivation of the Scale	24
	PART II. THE SPEED AND QUALITY OF HANDWRITING IN SEVEN SCHOOL SYSTEMS	
8.	Differences between Systems	29
9.	The Relation of Differences in Results to Differences in Means and Methods of Teaching Handwriting	32
10.	Differences between Individuals within the Same School System	33
11.	The Relation Amongst Individuals between Speed and Quality	35
12.	The Relation of the Quality of Slow Writing to the Quality of Rapid Writing by the Same Individual	36
13.	Miscellaneous Comments	36
14.	A Scale Based on Equally Often Noted Differences in Quality	40

SCALE A	between pp.	10 and 11
SCALE B	" "	24 and 25
SCALE C	" "	40 and 41

SCALE C

A SCALE FOR HANDWRITING OF CHILDREN IN GRADES 5 TO 8

Quality 18. Sample 135

showed that the rise and fall of the tides
The attraction of the moon and sun upon

Quality 17. Sample 132

Baltimore is one of the first cities
whose streets were lighted by gas.

Quality 16. Sample 58

stepped lightly into Warren's carriage and
held out a small card. John vanished behind
the bushes, and the carriage moved along

Quality 15. Sample 10

Then the carelessly dressed gentle-
man stepped lightly into Warrens
carriage and held out a small card.

Quality 14. Sample 1

Then the carelessly dressed gentle-
man stepped lightly into Warren's
carriage and held out a small
card. John vanished behind the

Quality 13. Sample 55

Then the carelessly dressed gentleman stepped
lightly into Warren's carriage, and held out a

Quality 12. Sample 52

Then the carelessly dressed gentleman stepped
lightly into Warren's carriage and held out a small
card, John vanished behind the bushes and the

Quality 11. Sample 44

the curb. Albert was sadly
striding across the field at times
now and then to look back at

Quality 10. Sample 34

an instant leaving only a poor old lady
on the curb. Albert was sadly striding across
the field stopping now and then to look
back at the village or staring at the bit of paper

Quality 9. Sample 51

Then the carelessly dressed gentle-
man stepped lightly into Warren's
carriage and held out a small card
John vanished behind the bushes

Quality 8. Sample 22

which had been gathering about them melted
away in an instant leaving only a poor
old lady on the curb. Albert was sadly
striding

Quality 7. Sample 126

card, John vanished behind the
bushes and the carriage moved

LB N '12

HANDWRITING

Section 1. Introduction

Handwriting may profitably be studied from three points of view:—that of the physiology and psychology of movement,[1] that of the part it may play in the intelligently directed activities of child life in schools,[2] and that of the direct examination of the quality and speed of handwriting secured by various forms of school training. But to any study of it there is one very desirable preliminary — some means of measuring the quality of a sample of handwriting.

At present we can do no better than estimate a handwriting as very bad, bad, good, very good, or extremely good, knowing only vaguely what we mean thereby, running the risk of shifting our standards with time, and only by chance meaning the same by a word as some other student of the facts means by it. We are in the condition in which students of temperature were before the discovery of the thermometer or any other scale for measuring temperature beyond the very hot, hot, warm, lukewarm, and the like, of subjective opinion. We opine roughly that, at a fairly rapid rate, writing-movements in which the forearm shares will produce a better quality of handwriting than movements confined more exclusively to the thumb and fingers, but no one could estimate with surety and precision *how much* better the best rapid " free-arm " writing is than the best equally rapid " finger-movement " writing. We opine roughly that drills in which good writing serves some end of consequence to the

[1] No attempt is made in this article to report any results of physiological or psychological analysis of the behavior involved in handwriting. The student of this aspect of the subject should consult especially the investigations of Preyer, Judd and Freeman.

[2] No attempt is made in this article to report the experiences or opinions of students of education with respect to the utilization of the original tendencies of children so as to secure a rational and economical cultivation of handwriting as an expressive art.

children will be more efficient than drills for mere penmanship, but no one could estimate *how much* more efficient they will be. We know that some schools secure better writing at a given speed than do other schools, but no one could tell *how much* better in any terms sure of understanding and agreement; for we have no scale to measure handwriting by. No pupil, teacher, or superintendent of schools knows how well any child, class, or group of children writes in anything approaching the sense in which we know how hot any liquid is or how long a wire is.

The main purposes of this monograph are to describe the means by which a graphometer or scale for handwriting may be made, to present such a scale for the handwriting of children in grades 5, 6, 7, and 8, to explain how such a scale is to be used, to present a similar scale for adult women's handwriting, and to mention some of the facts and questions of importance to which the discovery and use of these scales have led.

Many circumstances have combined to prevent me from giving at this time anything like a perfect scale. The individual differences amongst competent judges in rating any example are so great that to get for it a measure accurate within one per cent of the difference in merit between the best and the worst of grammar-school (i.e., grades 5 to 8) writing requires that at least 200 judges rate it. I have not been able to command the services of so many. For the greatest practical convenience a scale should have for any quality samples of all the common styles of children's writings, and should include about ten qualities differing each from the next by equal steps—equal, that is, within, say, four per cent of a step or one half of one per cent of the difference between the worst and the best grammar-school writing. But to get such samples one would need to have several thousand samples of each style of writing, and to have about half a million ratings made. This means roughly four thousand hours of labor. The final selections of samples for the scale should properly be made from very many printed reproductions such as will form the scale itself. The cost has prohibited me from making many of these.

The scale is presented now, in spite of its imperfections, for these reasons: It is the result of some twenty thousand ratings

and ensures measurements far more accurate than anyone could make without it. For the present needs of school practice and educational research, a very precise instrument for measuring handwriting is not required. The best way to get a more perfect scale is by the use of this one as a starting point.

This scale is then offered as a preliminary scale whose imperfections the maker is, perhaps, more conscious of than any critic will be. I beg the reader to bear this in mind, since, for the sake of simplicity in description in what follows, I shall not in each case state the fact that a quality or point on the scale is determined only to a certain approximation, and the fact that the differences between successive qualities are only approximately equal.

PART I

The Measurement of the Quality of Handwriting

Section 2. The Construction of a Scale for Quality of Handwritings of Children in Grades 5 to 8

If one selects from children's written work 1000 samples ranging from the best to the worst handwriting found in grades 5 to 8 and tries to rank these 1000 samples in order of merit for handwriting, one finds that he cannot make 1000 such ranks. Some of the handwritings will be indistinguishable in "goodness" or "quality" or "merit." Nor can one make 100 such ranks. Nor can one make 40. One can make about 20, but if he so ranks the samples a number of times he gets substantially the same average result as he gets when he ranks them a number of times in 10 or 11 groups. To get an individual's judgment of the relative merits of the 1000 samples it is sufficient to have him rank them in 10 or 11 groups three or four times. If he grades in 10 groups and tries to make the difference in "goodness" or "quality" or "merit" all equal,—to make, that is, the sample he puts in the highest group (call it 11) as much superior to those in the next highest group (call it 10) as the latter are to those he puts in the second from the highest group (call it 9), etc., etc.,—we have in the average[1] result of his groupings his judgment of the relative merits of the samples in a specially convenient form. For instance, if he grades sample 217 as in group 5 three times, as in group 4 once, and as in group 6 once, and grades sample 218 as in group 6 three times, in group 5 once, and in group 7 once, he judges 218 to be "1" better than 217, "1" being, in the individual's judgment, one tenth of the difference between group 1 and group 11.

If thirty or forty individuals chosen from competent judges of handwriting thus judge the 1000 samples, the average[1] of all

[1] Except for certain factors which will be described in section 7.

their gradings give approximately the relative merit of each sample in the judgment of competent judges in general. If they grade sample 317 in group 3 two times, in group 4 five times, in group 5 thirteen times, in group 6 thirteen times, in group 7 five times, and in group 8 two times, their average or median grade for it is 5.5. If their average or median grade for sample 318 is 6.4, they esteem 318 as .9 better than 317. The .9 means, in their judgment, nine tenths of one tenth of the difference between grade one and grade eleven.

If now from all the 1000 samples we could find some which were graded exactly 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11 by the average or median[1] judgment of 30 or 40 competent judges, each grading the set into groups 1 to 11 by what he thinks are equal steps in merit, we would have a very useful scale of merit in handwriting. It would include all grades from the worst to the best and would proceed by what were, by the average competent opinion, equal steps. Or if we could find some graded 1.5, 2.4, 3.3, 4.2, 5.1, 6.0, 6.9, 7.8, 8.7, 9.6, and 10.5 we would have a scale nearly as useful. It would not be so likely to include the very worst and very best samples, but would proceed by equal steps, as before.

The scale which I shall proceed to describe was obtained by a method in principle the same as the above.

Such a scale could be got in a different way, as follows: Suppose competent judges to compare each sample with every other, stating in each case which was better. If then we picked out samples *a, b, c, d,* etc., such that *a* was judged better than *b,* just as often as *b* was judged better than *c,* and just as often as *c* was judged better than *d,* and so on, we could have, in samples *a, b, c, d,* etc., a scale by equal steps, if two other conditions were fulfilled by them. The first of these conditions would be that *a* should not be judged better than *b* and worse than *b* *equally* often. For if it were, *a* would be equal to *b, b* to *c, c* to *d,* and so on, and we would have no extent to our scale. The second of these conditions would be that *a* should not *always* be judged better than *b.* For, if it were, it might be just enough better to barely be so judged, or it might be very, very much better.

[1] Except for certain factors which will be described in section 7.

Only if differences are not always noticed can we say that differences equally often noticed are equal. But if we had, as a result of the judgments, facts like those below, we could say that *a, b, c, d,* etc., represented samples of writing progressing by equal steps of difference in quality.

1000 comparisons of *a, b, c, d,* etc., being made:

a was judged better than *b* in 73 per cent., equal to *b* in 11 per cent., and worse than *b* in 16 per cent. of the judgments.

b was judged better than *c* in 73 per cent., equal to *c* in 11 per cent., and worse than *c* in 16 per cent. of the judgments.

c was judged better than *d* in 73 per cent., equal to *b* in 11 per cent., and worse than *b* in 16 per cent. of the judgments, and so on for *d-e, e-f,* .....*n.*

The scale which I shall describe was tested throughout by this second method. The two methods do not give results that correspond exactly. The variations follow this rule: Judges will notice differences between poor samples when they compare them directly one with another which they would not count in rating them by a mental scale. For example, suppose samples *a, b, c,* and *d* to be rated 10, 9, 3, and 2 by comparison with a mental scale of eleven grades by equal steps. The percentage of judges regarding 10 as better than 9 will be smaller than that regarding 3 as better than 2.

Since we get two different scales by the two methods, there are four alternatives. We may adopt one or the other or combine them, or give the results by both methods. I shall take the latter alternative, but shall at this point present only the scale as derived by the first method. In a later section (Section 12) the scale is derived by the second method will be presented.

The scale given here is then a scale in which the steps of difference are equal in the sense of being called equal by competent judges. Equal will mean just this in the next three sections. They are not equal in the sense of being equally often noticed when the single question "better or worse," is answered for each sample in connection with every other sample. The differences in the upper part of the scale would be less often so noticed than those in its lowest third.

Section 3. The Nature of the Scale

Scale A is the scale for merit of the handwriting of children of grades 5 to 8. It is not a scale of merit of the writings of children of grades 1 to 4 or of the writings of boys and girls of high-school age. It can, however, be more or less well used for such cases until we get more appropriate scales. Each set of samples represents a point on this scale.

The use of 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, and 17 for these qualities of handwriting means, first of all, that 14 is as much better than 13, as 13 is than 12; that 13 is as much better than 12, as 12 is than 11, and so on. In the second place it means that quality 14 is two times as far above 0 merit in handwriting as quality 7 is; that quality 16 is twice as far above 0 merit in handwriting as quality 8 is, and so on. Zero merit is defined roughly as writing as bad as sample 140 (see page 16), as a handwriting, recognizable as such, but of absolutely no merit as handwriting. The use of several samples under one quality means that those samples are of equal merit. The scale includes samples of as many different styles as could be obtained, so that in using the scale the merit of any sample of any style of writing can be quickly ascertained by comparison with the scale. The scale extends in actual samples by children from nearly the worst writing[1] of fourth-grade children (quality 5) to nearly the best writing of eighth-grade children (quality 17). Quality 7 is nearly the worst writing of fifth-grade children.

The scale includes a sample of a copy-book model which is rated by competent judges as of approximately quality 18, two samples of fourth-grade writing which are judged to be approximately of qualities 6 and 5, and a very bad writing, artificially produced, which is rated by competent judges as of approximately quality 4. The scale thus extends from a quality better than which no pupil is expected to produce, down to a quality so bad as to be intolerable, and probably almost never found, in school practice in the grammar grades.

If one had a finer scale, its use would give but slightly more accurate results, and would require more practice and more time.

[1] In a formal exercise in writing at their "natural" rate.

Any specimen of handwriting is measured by this scale by putting it alongside the scale, as it were, and seeing to what point on the scale it is nearest. Thus, the sample on page 9 (sample 9) is measured by comparing it with those of Scale A. I judge it to be between quality 15 and quality 14 and assign it the measure 14 rather than any other *unit* measure of the scale. If one wishes to measure more finely than to units, he can add or subtract a fraction according as the sample to be measured seems better or worse than the quality of the scale to which it is nearest.

The sample to be measured should, for convenience, be examined with the entire scale in view. If the scale's samples are arranged in order on a table or against a wall, the examined sample is easily compared with them. The measurer then decides what quality of the scale the sample possesses and records the measure. The measurer should be, of course, careful not to decide its grade because of its likeness in *style,* but only because of its likeness in *quality* to some sample of the scale. If, for instance, one has a pronounced vertical that is really of quality 7, one must not call it quality 8, because it is in style more like sample 14 than like the sample of quality 7. The measure may be made more and more accurate by having other judges also measure, each always in ignorance of the ratings given by the others. In default of other judges, the measure may be made more accurate by rating the sample two or more times, each time in ignorance of the ratings previously given. An individual may be measured more accurately by using several samples of his writing, each being rated in ignorance of the ratings given to the other sample.

Section 4. Criticisms of the Scale

The scale has, as previously noted, some defects. First of all, not all styles of writing are represented on the scale, much less at each point of quality on it. For example, we have no pronounced backhand writings of certain qualities and no very pronounced forward slant of certain qualities. There are hardly any markedly angular writings on the scale. This defect can be at any time remedied by securing enough samples of children's writing of the missing sorts at approximately the qualities in question,

Sample 9

carriage and held out a small cord.
John vanished behind the bushes, and
the carriage moved along down

selecting, with the aid of thirty or forty competent judges, samples whose merit is exactly 8 or 10 or 12 or 14 as the case may be, and adding these to the scale. I shall be grateful to any one who sends me collections of children's handwritings of styles not represented in the scale.

Each such sample should be accompanied by a statement of all the grades assigned to it on our scale by at least ten or twelve competent observers, each of whom measures it with the scale and rates it in complete ignorance of the ratings given by all the other judges. It is desirable, though not necessary, that the writings be on unruled paper.

In the second place, the qualities below 5 and above 17 should perhaps be represented in the scale by actual children's writings. This defect could be remedied by collecting children's handwritings that were superlatively bad and superlatively good. I shall be grateful to anyone who sends me samples of children's writing which are notably better than quality 17 or notably worse than quality 5.

In the third place, although I have so far spoken of the qualities 5, 6, 7, 8, 9, 10, etc., as if they might be absolutely these amounts—as if the 13's might be all absolutely equal in merit and all absolutely halfway between any one of the 12's and any one of the 14's—this is not exactly the case. As was noted on page 3, the scale is only approximate. 16 on the scale does not pretend to mean 16.00000, but between 15.9 and 16.1. 8 does not pretend to mean 8.0, but between 7.9 and 8.1. And as a matter of fact, although I have had a thousand samples graded and have chosen as wisely as I could, some of the samples do vary in merit from 7, 8, 9, 10, etc., by more than .1 plus or minus. Even after one has picked samples that vary only that much, the relations may be altered in the process of making the electrotypes from which the scale is printed or in the process of printing itself. This defect can be remedied by the expenditure of enough time and money in getting more samples, having them graded by more judges, reproducing more of them in electrotypes, and having these reproductions graded again by more judges. In this work I am now engaged. The defect is, however, of little consequence to any use to which any of my readers is likely to put the scale.

SCALE A
A SCALE FOR HANDWRITING OF CHILDREN IN GRADES 5 TO 8

The unit of the slope equal upper loop, one-half of the distance between the head and waist of the normal writing of 10-year-old children in grade 5. The difference between the upper loop and the lower loop of the combined unit is equal to the distance between the top of the capital letter and the top of the lower loop.

Quality 18. Sample 125

showed that the rise and fall of the tides
the attraction of the moon and sun upon

Quality 17. Sample 141

Then the carelessly dressed gentleman
stepped lightly into Warren's carriage and
held out a small card, John vanished be

Quality 16. Samples 32 and 84

Then the carelessly dressed gentle-
man stepped lightly into Warren's
carriage and held out a small

gentleman stepped lightly into
Warren's carriage and held out
a small card, John vanished

Quality 15. Samples 49 and 89

lightly into Warren's carriage and held out a
small card, John vanished behind the bushes
and the carriage moved along down the drive

held out a small card, John vanished behind
the bushes and the carriage moved along
down the driveway The audience of passers-

Quality 15. Sample 47 and 90

John vanished behind the bushes
and the carriage moved along
down the driveway. The audience

Then the carelessly dressed gentle-
man stepped lightly into Warrens
carriage and held out a small white

Quality 14. Samples 54 and 19

Then the carelessly dressed gentleman
stepped lightly into Warren's carriage
and held out a small card, J

Then the carelessly dressed gentleman
stepped lightly into Warrens carriage and
held out a small card, John vanished behind

Quality 13 Samples 55, 24 and 26

Then the carelessly dressed gentleman stepped
lightly into Warren's carriage and held out a

vished behind the bushes and the car-
riage moved along down the driveway.
The audience of passers-by which had

Then the carelessly dressed gentleman
stepped lightly into Warren's carriage and

Quality 13. Sample 4

Then the carelessly dressed
gentlemen stepped lightly
into Warren's carriage and

Quality 12. Sample 30

lightly into Warrens carriage and held out a
small card, John vanished behind the bushes
and the carriage moved along down the drive

SCALE A (Continued)

Quality 12. Samples 7 and 32

behind the bushes and the carriage moved along down the driveway. The audience of passers-by

Then the carelessly dressed gentleman stepped lightly into Warren's carriage and held out a small card, John vanished behind the bushes and the

Quality 11. Samples 23 and 45

stage moved along down the driveway. The audience of passers-by which had been gathering about them melted away

along the down the driveway. The audience of passers-by which had been gathering about them

Quality 11, Sample 206

John vanished behind the bushes and the carriage moved along down the driveway. The audience

Quality 10. Sample 17

driveway. The audience of passers-by which had been gathering about them melted away in an instant leaving only a poor old lady on the curb. Albert was sadly thinking

Quality 9. Samples 31, 21 and 28

Then the carelessly dressed gentlemen stepped lightly into Warrens carriage and held out a small card, John vanished behind the

by which had been gathering about them melted awayin an instant leaving only a poor old lady on the curb. Albert was sadly

Then the carelessly dressed gentleman stepped lightly into Warren's carriage moved and held out a small card, John vanished

Quality 8. Samples 14 and 46

moved along down the driveway. The audience of passers by which had been gathering about them melted away.

Then the carelessly gentleman stepped lightly into Warrens carriage and held out a small card, John vanished behind the bushes and the carriage moved

Quality 7. Sample 126

card, John vanished behind the bushes and the carriage moved

Quality 6. Sample 12

gathering about them melted away in an instant leaving only a poor old lady.

Quality 5. Sample 6

bushes and the carriage moved along down the driveway. The audie

Quality 4. Sample 121

seated on the curb like my dreams and

For the variations in the scale are trivial compared to the variations in individual judgment. I have measured the quality of each sample in the scale to tenths of a step, subject to slight changes had more judges been available, and apart from variations in the printing. For example, the quality of sample 49 in the scale is 15.1, not exactly 15.0.

Similar figures for each sample in the scale are given below. If any one wishes to have the values of each sample as precise as possible he may substitute these values. In scientific studies of handwriting in schools this should be done, but in practical grading by teachers the 5, 6, 7, 8, etc., of the scale may be kept without the decimal alterations.

What changes might be made in the qualities, if the consensus of thousands of judges were to replace the consensus of from twenty to seventy, is shown in the figures in the third column, which give the probable average divergences of the former consensus from the latter. They show that the scale is not nearly so precise as, say, a 25 cent scale for weight. But, on the other hand, the superiority of the scale to the personal opinion of any one teacher or investigator is enormous. The latter would have a probable average divergence of from 1.0 to 1.6 from the consensus of a thousand competent judges.

TABLE I

Sample	Quality	Probable average divergence of the estimated quality from an estimate by an infinite number of judges
32	16.1	.14
84	16.2	.43
47	15.0	.19
49	15.1	.18
89	15.0	.39
90	15.1	.35
19	14.0	.20
54	14.0	.19
4	12.9	.20
24	13.1	.18
26	12.9	.18
55	13.1	.21
30	11.9	.19
7	12.0	.20
52	12.0	.20
23	11.0	.20

Sample	Quality	Probable average divergence of the estimated quality from an estimate by an infinite number of judges
45	11.0	.19
106	11.0	.28
17	10.2	.18
21	9.1	.15
28	8.9	.15
31	8.9	.14
48	8.0	.14
14	8.1	.19
126	7.0	.40

The reader, in examining the scale, may think that some of the samples called equal are really unequal. If he objects to vertical writing, he may, for instance, think that sample 55 in Scale A is at least one step worse than sample 24. Such criticisms of the scale are, however, really strong arguments in its favor. For such a critic is surely wrong. That he denies the correctness of the average opinion of forty competent judges means simply that his own judgment is partial or crude, and the fact that each individual's judgments of handwriting are thus partial and crude proves that he needs a scale representing the general judgment of competent people to help him to judge and to teach him to eliminate the unfairness in his own future judgments. If no one felt any disagreement with this scale, it would not be so valuable as it is under the condition that many individuals will think it wrong. For those who are unfair to any style of handwriting or who overemphasize beauty in comparison with legibility, or evenness in comparison with "character," or the reverse, can be proved by the scale to be unfair—that is, to diverge from the average judgment of competent people in general. If they are intelligent, they can learn from the scale to correct their bias.

It is possible, however, that some critic may deny the value of the average judgment of competent people in general and declare that though that judgment pronounces two handwritings equal in merit, *he* knows that they are not equal. Now conceivably he might be right. But the chances are enormously against his being right, and science naturally cannot count his assurance as of more weight than that of any other judge of equal competence.

Some more sophisticated critic may object, not that he knows that *this* scale is wrong and prefers his own supposed competence to that of forty of his peers, but that no one can know whether this or any such scale is right. For, he will add, any such scale is subjective,[1] representing only what certain individuals think about the merit or value of samples of handwriting. In this there is some truth. There is no value in average opinion as such. The world was as round, when the most competent judges thought it flat, as it is today. If it should some time be proved that evenness of width of line was the sole criterion of real merit in handwriting, the scale would be wrong. But in the case of handwriting the only available criterion of real "merit" or "quality" or "goodness" is the average judgment of competent people. A hundred years from now merit in handwriting may mean something different from what it now means and a new scale may be required. But what it then means will then be determined by the average judgment of competent men and shown in a scale derived just as this one has been derived. What merit *does now mean* is precisely the thing measured by this scale. Merit in handwriting in the judgment of competent people today is the composite of qualities, each duly weighted, wherein the samples marked 12 are as much better than the samples marked 10 as the latter are than those marked 8, etc. The scale measures not only some absolute merit, but merit as now defined in the average judgment of forty or more persons chosen at random from the competent. And no other sort of merit is so well fitted to be the basis of a scale.

A far more sagacious criticism than either of these would be that a scale like this for merit in general is less useful than a scale for legibility alone, or for beauty alone, or for "character" alone, or for ease alone. Of course, I admit that such specialized scales are highly desirable, and I hope that this scale for general

[1] If this report were addressed to students specially interested in logic and scientific methods applied to the social sciences, it would be worth while to show here that the objectivity of a scale for length as compared with the subjectivity of a scale for merit of handwritings, or moral worth of acts, or beauty of poems, means only a closer likeness amongst men in their judgments, not a radically different sort of judgment. Being far, far more alike in sense-organs and muscles than in the central connections of neurones, we agree far better in comparing lines and weights than in comparing handwritings or poems.

merit will stimulate others to the labor of making similar scales for legibility alone, beauty alone, and so on. But it seems sure that the scale of most importance and usefulness is that for general merit. General merit is that for which school grades are oftenest given, in respect to which school systems or classes are oftenest compared, and with which other features of a pupil's achievements are oftenest related. Moreover, only after a scale for general merit has been made can one measure the extent to which legibility, beauty, etc., respectively determine general merit.

So much for criticism of the general method of constructing the scale. I turn now to possible criticisms of the scaling itself.

Some one may ask why 4, 5, 6, 7, 8, etc., are used as the values of the samples of Scale A instead of some other equal-step series of numbers such as 1, 2, 3, 4, 5, 6, etc., or 65, 67½, 75, 77½, 80, etc. The step is made 1 rather than 2½ because one cannot judge samples precisely enough to profit by more than 18 divisions in a scale. Hence the time spent in deciding whether to call a sample measured by the scale 78 or 76 or 77 and in later computing with the large numbers would be largely wasted. The ratio of the highest to the lowest children's writing in the scale is made 17 to 5 (or roughly 3½ to 1), instead of 6 to 1 or 13 to 9 (97½ to 67½) because, from the average opinion of competent judges and the facts of individual differences in motor ability, zeal for handwriting, and other factors determining the quality of a pupil's writing, the best writing from children in these grades seems likely to possess less than six times as much merit as the worst, but more than one and a third times as much—in other words, to be less than six times as far, but more than twice as far, beyond zero merit.

That is, the scale was arranged so that the numbers representing the distances beyond zero of the best and worst samples of children's writing in our scale should stand in the ratio of approximately 3½ to 1, and also so that the numbers on the scale should be the smallest compatible with as accurate measurement of handwritings as educational theory and practice need. If any one prefers as a scaling 15, 17, 19, 21.....43, or 3, 4, 5, 6 ....17, or 7, 8, 9, 10....21, it would be hard to prove to him that his choice was inferior to the 4, 5, 6, 7, 8....18 used. The

essential thing is that the steps be equal, and that the ratio which the amount attached to the best children's writing bears to the amount attached to the worst be a reasonable one.

Having defined what was meant by 0 merit (see sample 140 on page 16, I judged as best I could the distance of sample 141[1] therefrom in terms of the distance of sample 2[2] therefrom. The judgment of 3 2/5 times is by no means final. Indeed I am now engaged in an investigation aiming to revise it. I could argue plausibly for a ratio as low as 2½ to 1 or for one as high as 5 to 1. But a ratio somewhere between 3 to 1 and 3½ to 1 seems the most reasonable.

The whole matter of the choice of an absolute 0 for merit in handwriting, and of the consequent absolute values of the points on the scale, is one involving many intricate considerations out of place in this discussion. I fear that in touching upon it at all I may have perplexed some readers. Such may rest confident that in using the 4, 5, 6, 7, 8, 9, 10, 11, 12, etc., of the scale in measuring a sample of handwriting as they would use 4, 5, 6, 7, 8, 9, 10, 11, 12, etc., dollars in measuring the value of a book or a jewel or a trunk, they will commit no error of much consequence or, at least, no error so great as they would be likely to commit by measuring it in any other one way.

Another criticism may be that the scale does not guarantee agreement among the observers using it to measure a sample of handwriting. The same sample may, it will be said, be measured by one person as equal in merit to 8, by another as equal to 10, and by still another as equal to 9. This is true, but it is not the fault of the scale. Observers will disagree in their measurements made with the scale, but *not nearly so much as in measurements made without it.* No scale guarantees absolute agreement. Observers measuring the length of this line ———— to tenths of a millimeter will not agree. But they will agree better than they would if they had no scale and judged its length as a savage might.

[1] Nearly the best sample from children in grades 7 and 8.
[2] Nearly the worst sample from children in grades 4 and 5.

Sample 140, representing zero merit in handwriting. Zero merit is arbitrarily defined as that of a handwriting, recognizable as such, but yet not legible at all and possessed of no beauty.

Section 5. The Uses of the Scale

The topic of this section is fitly treated in the one statement: Any measurement of the quality of handwriting may be made more accurately and conveniently with the scale, either actually present or held in memory, than without it. The reader may apply this statement to whatever cases his interests suggest. I shall mention a few of the commoner uses and explain the function of the scale as a standard held in memory.

The class-room teacher has to measure the quality of a single pupil's handwriting in order to assign him a rating in comparison with his fellows and, better still, in comparison with his own past performances. If she uses the scale either by giving its numerical measures outright or by letting her A, B, C's, or 75, 80, 82, etc., per cents, or excellents, goods, fairs, etc., mean certain points on the scale, her ratings will have a definite meaning to the pupil, can have the same meanings that similar ratings by other teachers in the school have, and may be used to measure the actual improvement of the pupil month by month and year by year. She can more easily and more accurately measure the relative values of the different methods of teaching which she may from time to time employ, of different lengths of periods for drill, and the like.

A principal or supervisor or superintendent of schools needs to measure the quality of the handwriting of individuals, of classes, and of all classes of the same grade, in a school or system. If he has such measures honestly made by the scale, he can compare the work of one teacher with that of another, the work within his own school or city with that of other schools or cities and with that of his own city five years later, the work of schools using one system of writing with that of schools using other systems, and the like. If he tried without the scale to estimate the superiority or inferiority in handwriting of twelve-year-olds in city A to twelve-year-olds in city B, he would have to collect many samples in both cities and have them graded alike. He could define the amount of difference found only by actually exhibiting it in samples *or by making out a scale like ours,* defining it as I have done, and expressing the difference as such a distance on the scale. With the scale in use in both cities, on the

contrary, if marks are honestly given by the teachers, the superiority or inferiority of any group will be measured by the difference in the scale-values of the marks themselves.

The scientific student of education will use the scale to measure the quality of samples of handwriting from individuals, classes, cities, groups chosen for grade, age, sex, method of teaching, or length of time devoted to writing, and from any other sources. He will also be able to use any marks or ratings honestly given by teachers or others.

Whoever has any occasion to define a standard of quality in handwriting can define it unmistakably and conveniently by the scale. Business men can decide what quality they wish the schools to secure in the boy fourteen years old who is to apply for clerical positions. A supervisor can inform all the teachers of, say, grade 7 that the minimum requirement is, say, quality 11, at a rate of 50 letters per minute, that the average pupil must be made to write at quality 13 at a rate of 60 letters per minute, and so on. Whatever standard is set will be absolutely defined by those who set it and will be clear to all those who are to follow it.

The pupil himself may profitably know and use the scale. He may see by it what is expected of him and may tell how nearly he reaches the standard and how much he has gained.

Even if precision is not desired in the estimate of the quality of handwriting,—even if *good* and *bad* or *satisfactory* and *unsatisfactory* are the only ratings to be given,—the scale is none the less useful. For if *good* and *bad,* or *satisfactory* and *unsatisfactory* are to mean anything, they must mean handwritings above and below some point on some scale of merit. They can be properly defined only by locating that point. And until some better scale is available that point can be located only by exhibiting samples or by stating the numerical value these samples would have on our scale.

To put the whole matter in a word, any measurement of the quality of handwriting should be made by the scale and reported in terms of the scale, for substantially the same reasons that any measurement of the length of an object should be made with a linear scale and reported in meters or feet.

Measurements may be made by the scale without the comparison of the sample with the actual scale itself. Just as one uses his experience of feet and inches as a mental standard whereby he estimates more or less accurately the length of pencils, tables, windows, and the like, without an actual ruler or tape, so one may come to estimate that this sample of writing is about quality 16, that one about quality 9, and the like, from the mental standard left from examination and use of the actual scale. The scale should always be present for reference in any measurement which requires exactitude, but it will do its greatest amount of service, not by actually serving as a foot-rule for quality in handwriting, but by creating in the minds of teachers mental standards to be used in even the most casual ratings of everyday school-room life. To one who uses the scale quality 18 or quality 15 or quality 7 comes to be a definite agent in determining all judgments, just as 18 inches or 15 pounds or 7 dollars is. Just as a child learns to think about length correctly and with fair precision without a ruler in his hand, by having measured off lengths with it, so teachers may come to think about handwriting correctly and with fair precision without the scale before their eyes by having measured handwritings with it. Just as the thermometer teaches us to supplant the vague "very cold," "cold," "moderate," "warm," "hot," and "very hot" by "about 0," "about freezing," "about 60 degrees," "about 70 degrees," "about 80 degrees," "nearly 100," and the like, so the graphometer can teach us to supplant the vague "illegible," "very hard to read," "a good plain hand," and the like, by judgments which mean something definite and constant to those who make and those who hear them.

Section 6. A Scale for Quality in Adult Women's Handwriting

The scale for adult women's handwriting consists of only six points, each represented by only one sample. Let us call these samples *a, b, c, d, e,* and *f*. They represent the best selection that I could make of writings ranging from nearly the best to nearly the worst of the ordinary writings of some five hundred women teachers and students and differing progressively by

equal degrees of merit. The derivation of the scale was as follows:

Thirty judges rated samples *a, b, c, d, e,* and *f,* together with from 37 to 456 other samples. The ratings given were from 1 (the lowest grade) to 11 (the highest), grades 1 to 11 being roughly shown by samples and the requirements being made that the grades 2, 3, 4, 5, etc., should represent grades of merit differing by equal steps. The number of the samples was reduced from 456 to 37 by gradually dropping samples which seemed unlikely to be near the points 1 to 11. The result of the thirty ratings is shown in Table II.

TABLE II

THE QUALITY OF SAMPLES *a, b, c,* ETC., AS MEASURED BY 30 JUDGES FROM THE ORIGINAL WRITINGS

Quality	Frequencies of Each Quality for Each Sample					
	a	*b*	*c*	*d*	*e*	*f*
1	20	1	1			
2	7	12				
3	1	6	2	1		
4	2	6	7	1		
5		2	10	6	1	
6		1	5	3	2	
7			3	8	6	2
8		1	1	3	4	
9			1	5	6	3
10				1	6	6
11				3	5	19

Bearing in mind that a rating of quality 1 means 1 *or worse than* 1 and that a rating of quality 11 means 11 *or better than* 11, it is clear that in the combined judgment of all 30 judges *a, b, c, d, e,* and *f* represent qualities progressing by approximately equal steps. Thus 10 of the judges ranked *a* as better than 1, 10 ranked *b* as better than 3, and 10 ranked *c* as better than 5, 12 ranked *d* as better than 7, and 11 ranked *e* as better than 9. Of the 20 judgments of *a* as 1, it is probable that about 10 would have been "worse than 1" had the series included a lower range. Of the 18 judgments of *f* as 11, it is probable that about 10 would have been "better than 11" had the series included a higher range. The median values of *a, b, c,* etc., with this interpretation of the grades 1 and 11, are: 1.0, 2.833, 5.0,

7.125, 8.833 and 10.94, the differences in quality being respectively 1.833, 2.167, 2.125, 1.708, and 2.107.

These six samples were then printed and were graded in their printed form, together with seven other samples of approximately the qualities 1, 3, 5, 5, 7, 9 and 11, by thirty-eight judges. The ratings in this case were in 6 grades, to progress by equal steps. These were called by the judges 1, 2, 3, 4, 5 and 6, but represent respectively 1, 3, 5, 7, 9 and 11 of the gradings just presented in Table II. Hence in Table III, which gives the results of the gradings by these thirty-eight judges, I shall use 1, 3, 5, etc., for 1, 2, 3, etc.

TABLE III

THE QUALITIES OF SAMPLES *a, b, c,* ETC., AS MEASURED BY 38 JUDGES FROM THE PRINTED REPRODUCTIONS

Quality	Frequencies of Each Quality for Each Sample					
	a	*b**	*c*	*d*	*e*	*f*
1	32	7				
3	5	21	8	3		
5	1	9	19	12	1	
7			9	19	6	3
9			2	3	22	13
11				1	9	22

* Only 37 judges rated this sample.

The median ratings for *a, b, c,* etc., are .8, 3.1, 5.1, 6.4, 9.1, 10.7.

These thirteen printed samples were then rated together with from 58 to 104 samples of children's handwriting, including samples much better than the best of the adults', by 26 judges. The ratings were from 1 to 11, but the meanings of these numbers were unlike those attached to them in Tables II and III, except in the case of the 1. The 3, 5, 7, 9, and 11 of Tables II and III have approximately the values 2.4, 3.8, 5.2, 6.6, and 8. Finally, the thirteen samples were rated, together with 120 samples of children's writings, including some still better and some still worse, by 9 judges. The ratings were 0 to 12 but the values of the 1, 3, 5, 7, 9, and 11 of Tables II and III were, as before, approximately 1, 2.4, 3.8, 5.2, 6.6 and 8. The median values attached by the 35 judges were, for *a, b, c, d, e,* and *f,* in order, 1, 2.4, 3.83, 5.3, 6.5, and 7.9.

We have then as a result of the three series of judgments, numbering 103 in all, the following:

Differences between *a* and *b*, *b* and *c*, *c* and *d*, etc.:

I. Using the median ratings of 30 judges (ink samples): 1.83, 2.17, 2.13, 1.71, 2.11.

II. Using the median ratings of 38 judges (print): 2.3, 2.0, 1.3, 2.7, 1.6.

III. Using the median ratings of 35 judges (print, long series) reduced to equivalences with (I) and (II): 2.0, 2.04, 2.1, 1.7, 2.0.

Average differences: a-b, 2.04; b-c, 2.07; c-d, 1.84; d-e, 2.04; e-f, 1.91.

The approximate equality of the steps may be verified by ascertaining how often *b* is rated higher than *a*, how often *c* is rated higher than *b*, etc., that is, by an adaptation of the so-called method of right and wrong cases. The facts are as follows:

TABLE IV

COMPARISONS OF *a*, *b*, *c*, *d*, *e*, AND *f* BY 102 JUDGES

	Long series, written samples	Series of 13 printed samples	Series of 71 to 133 printed samples	All series together
No. of comparisons	30	37	35	102
b rated as better than *a*	25*	26	23	74
c rated as better than *b*	26*	25*	27*	78
d rated as better than *c*	23	25*	25*	73
e rated as better than *d*	23	29	19	71
f rated as better than *e*	23*	24	26	73

In the starred cases the obtained figure was 1 less than that printed, but the number of comparisons it was from was also 1 less than that printed at the top of the column.

Samples *a*, *b*, *c*, *d*, *e*, and *f* thus represent points on a scale of quality differing each from the next by approximately equal steps. We can properly call their values in order x, $x+2$, $x+4$, $x+6$, $x+8$, and $x+10$ where 1.0 equals a difference roughly equal to one-tenth of the difference between the best ten and the worst ten of a thousand samples each from an adult woman student and x equals the average quality of the worst ten of the thousand. To be more precise we should call them, in order, x, $x+2.0$, $x+4.1$, $x+6.0$, $x+8.0$, and $x+9.9$.

To turn these values into numbers referring to zero merit as a starting point we must define zero merit for adult handwriting and measure the distance of *x* from it.

This I have not attempted to do at all adequately since the need of an elaborate scale is not nearly so great in the case of adult handwriting as in the case of children's writing. Quality *x* of the adult scale is judged by the average of some forty individuals to be approximately equal to quality 8 of the children's scale. A difference of 1.0 along the adult scale is judged to be approximately equal to a difference of .7 along the children's scale. If we take the zero point for adults as approximately the same as for children of grades 5, 6, 7, and 8, the qualities of *a, b, c,* etc., may be taken as approximately equal, in order, to 8, 9.4, 10.8, 12.3, 13.6, and 14.9 or 15 on an absolute scale whose zero is a writing recognizable as an attempt to write, but of zero merit. Such a numbering would not be far wrong.

This adult scale very much needs samples of other styles at each point. Perhaps I should have delayed printing it until such had been obtained, but the labor and expense of collecting and selecting, by grading and gradual elimination, samples to fit exactly certain places on the scale is very great. The present scale has required thousands of gradings. It will be of great value in economizing the time and money of any one who wishes to make a better scale, if in no other way.

As a matter of fact, in spite of its lack of samples of all styles at each point, it will also be of service in every case where the quality of a woman's handwriting is to be definitely known.

For example, (1) the authorities of a college or a normal school wish to set a clear standard as to how good handwriting must be in order to make an examination paper, or a composition, or other written work, acceptable. If they set this standard as " at least as good as quality *c* of the Thorndike scale " every student, every member of the teaching staff, the faculties of other colleges, and the public can tell just what the standard is. There can be real as well as " paper " uniformity in the standard.

(2) In civil service examinations, examinations for teacher's licenses and the like, the standard of a certain quality by the scale at a certain minimum speed can be set and the candidates

can be exactly, impartially, and uniformly (all over the country, if desired) rated.

(3) The relation between (a) ability in handwriting under the pressure of school drill to (b) ability in handwriting in later life requires for study some adult scale. So also with any other relation of the quality of adult handwriting to anything.

I shall be indebted to any one who will send me samples of adult women's handwriting, especially of vertical writing of qualities *d, c, b, a,* and worse, of pronounced slant writing of qualities *d, e, f,* and better, and of pronounced backhand writing of all qualities. Each such sample should be accompanied by a statements of all the grades assigned to it on our scale by ten or twelve competent observers, each of whom judges in entire ignorance of the judgments made by all the others. It is desirable, though not necessary, that the writings be on unruled paper.

Section 7. The Derivation of the Scales[1]

Certain partial descriptions of the means and methods by which the children's scale and adult women's scale were derived have been given in sections 2 and 6. A full account of the derivation of either is inadvisable both because it would necessarily be extremely long and because much of the work done was such as I now know, from the very experience of doing it and seeing its results, to have been unnecessary.

I shall therefore give only such notes as are likely to be helpful to any one who is stimulated by this scale to construct similar scales for other educational products.

To construct a scale by which to measure various qualities (that is, amounts of merit) in handwriting ranging from, say, x to $x + y$, it is desirable to have samples of qualities, not only of every degree from x to $x + y$, but also of qualities worse than x and of qualities better than $x + y$. The reason is that otherwise the exact values of samples at x or x plus a slight amount and of samples at $x + y$ or $x + y$ minus a slight amount cannot be directly measured, but only inferred.

[1] The reader uninterested in educational measurements is advised to skip this section, and to turn at once to the more immediately practical discussion of differences amongst school systems with respect to speed and quality of handwriting.

SCALE B

A SCALE FOR HANDWRITING OF ADULT WOMEN

Sample f. Quality x = 10 (9.9). Valued as 15 (14.9) of Children's Scale

Exercises, p. 107. Questions 1–5. March 22-07.

1. Clearness is always desirable as a feature of the stimuli given by a teacher.
2. Novelty and pleasure lose their effect with use.
3. Expectedness and intensity offer an argument for stating the aim of each lesson.

Sample e. Quality x = 8. Valued as 13.6 of Children's Scale

Tends to impress its truth upon him

XVIII Get weights and some cheap substance & let them weigh out certain quantities and measure them. Let them exchange foreign for U. S. money.

XXIV Show water in its various forms. steam, ice, etc. By a chemical laboratory in which to make ex-

Sample d. Quality x = 6. Valued as 12.3 of Children's Scale

4. (1) Physiological, function: sneezing.
(2) Expression: wrinkling forehead, curling toes in, curling fingers in, raising eyebrows
(3) Effect: lifting arm, coughing, walking, pushing.

Sample c. Quality x = 4 (4.1). Valued as 10.8 of Children's Scale

whole and forming judgments that thinking is carried on — and for their mental furnaces it is necessary to hold ideas in mind. Such an individual would have no stability of character. as the state of mind determines to some extent the sensation and perceptions received. Therefore, the reaction from them without permanency of ideas a person under similar conditions would react very differently at different times.

Sample b. Quality x = 2. Valued as 9.4 of Children's Scale

Repetition frequently and near together
Result. Learn to shoot. Law of frequency, recency and resultant satisfaction - main factor
4 - a - 1 Instinct to put things into mouth - toys, lump of sugar. etc. discovered sugar was sweet so agreeable - formed habit of putting candy in mouth whenever it could reach it -
b - Instinct of puppy to chew. shoes, bones etc

Sample a. Quality x. Valued as 8 of Children's Scale

his past experiences would be of no service to him. ~~of the should~~ The fact that a dose of poison made him sick one day would not keep him from taking it again the next. He would never be able to make any preparations for the future, would be without imagination and inventive power.

For example, call x 1 and y 10. $X + y$ then being 11, x or 1 is nearly the worst and $x + y$ or 11 is nearly the best of a series of samples, ranging continuously from x to $x + y$.

If now any one is required to fix in mind 11 points including x (or 1) and $x + y$ (or 11) differing each from the next by equal amounts, and to rate each of the samples as 1, 2, 3, ——9, 10, or 11, according to which of these mentally fixed points it seems most like, he can err by rating a sample as 2 or 3 when it is really 1, but cannot err by rating it 0 or minus 1 when it is really 1. Similarly he can err by rating it 9 or 10 when it is really 11, but cannot err by rating it 12 or 13. For a sample really close to point 11, rated in the way just described by 33 judges, the results were:

Rated as 11 by 21 judges
Rated as 12 by 7 judges
Rated as 9 by 3 judges
Rated as 8 by 1 judge
Rated as 7 by 1 judge

The apparent average rating would then be 10.4 and the apparent median rating 10.7. When, however, the samples are increased by some of the real quality $x + y + 1$ (or 12) and the ratings are to be made at twelve points including $x + y + 1$ (or 12), a certain proportion of the judges rank the sample in question 12 and the average and median are raised to nearly 11.

Unless the set of samples to be rated includes some samples one, two, three, and even four grades better than the best quality $(x + y)$ to be represented in the final scale and also some samples one, two, and three grades worse than the worst quality (x) to be represented in the final scale, one cannot get the values of $x + y$ and x themselves save by inference.

Hence, to make a scale for the handwritings of, say, 10-year-old school children conveniently, it is necessary to have a collection of samples varying in quality from much below the worst to much above the best of their writings. This involves the use of "unnatural" samples, which may seem very objectionable, but which as a matter of fact does little or no harm.

In the case of a scale for the merit of English compositions by high-school pupils one should start from a collection of com-

positions ranging by small gradations from compositions much worse than the worse point on the final scale is to be, to compositions much better than the best point on the final scale is to be. Here the extremely bad ones may be obtained by artificial construction, from the feeble-minded, or from very old and stupid grammar-school children. The extremely good ones may be obtained from the printed or manuscript compositions in youth by gifted authors.

To get samples exactly situated at points differing progressively by equal steps requires that the original set range from one extreme to the other by very slight gradations. This means for practical purposes that one must have at the start a very large number of samples. After these have been graded by enough judges to rate each roughly, only those which are near the points to be represented by the scale need be graded further. As the value of each sample of this narrower selection is determined more exactly by further judgments, only those very near the points to be represented on the final scale need be preserved for still further judgments; and so on till the values of enough samples are determined to the degree of precision required for the scale itself.

Points on the scale exactly determined, but not at progressively equal steps, can be got with far less labor. If, for example, after a single rating I had picked samples at intervals from the best to the worst and then had only these few samples rated by the twenty to seventy judges, the value of each could have been stated nearly as exactly as is the case in the samples of the scale. But they would form a series like 17.33, 16.65, 16.28, 15.82, 15.40, 15.47, 15.23, 14.95, 14.7, etc., instead of the approximate 17, 16, 15, 15, 15, 15, 15, 14, 13, 13, 13, etc., of the scale. They would have served the purpose of a scale as well so far as aiding an observer to make exact measurements which any other observer could verify, and to report them unambiguously, but the labor of allowing for the decimal values or of computing measures expressed in awkwardly long numbers would burden each person using the scale. If the scale were designed for use only by scientific investigators of education, I should have economized in respect to the number of samples rated, had far more ratings of each sample, and presented a scale of very exactly determined

qualities but at irregular intervals. For the common use of pupils, teachers, and supervisory officers a less precise scale by approximately equal steps seemed far more valuable. Also the precise evaluation of each sample can be determined by many students each spending independently a little effort in getting the samples which I print rated; whereas the selection of samples varying by equal steps can be managed best under one individual's supervision.

It is possible that the determination of the amount of difference between two samples by the percentage of judges noticing the difference is preferable to the determination by the amount of difference between their median values as given by judges attempting to apply to each a scale of mentally equal differences. I used both methods. Experience of their use provides many facts of importance to methods of quantitative work in both psychology and education, but the facts would be of interest to only the small proportion of readers to whom surfaces of frequency of errors in judgment are familiar and esteemed friends.

In general, the experience in constructing this scale gives great encouragement to the hope that for many educational facts, units and scales may be invented that shall enable us to think quantitatively in somewhat the same way that we can about facts of physics, chemistry, or economics. It has been commonly supposed that the great complexity of such facts as examination papers in spelling, manifestations of interest in history, acts of moral significance, habits of industry, essays, poems, inventions, replies to questions demanding logical inferences, and other like results of education, prevents the samples composing any one such group from being measured by any one linear scale at all comparable to a foot rule or thermometer or galvanometer.

It is true that some judges find it hard to judge handwriting for the complex of legibility, beauty, ease, " character," etc., into which " quality " or " goodness " or " merit " resolves itself. But none of them found it impossible to do so, and most of them rated the writing for the complex,— " merit or goodness in your opinion,"—as readily as an appraiser would rank articles of sale by money price, or as a little child would arrange pieces of paper in the order of their size regardless of the fact that some were squares, some circles and some triangles.

The entire history of the judgments of the merit of handwritings supports the claim that if a number of facts are known to vary in the amount of any thing which can be thought of, they can be measured in respect to it. Otherwise, I may add, we would not know that they varied in it. Wherever we now properly use any comparative, we can by ingenuity learn to use defined points on a scale.

PART II

The Speed and Quality of Handwriting in Seven School Systems

The conclusions to be reported in sections 8 to 13 are based upon about 3000 samples of handwriting made in a formal test conducted by Dr. C. W. Stone[1] in seven school systems, five public and two private. These samples were scored in about 700 cases by six judges using no scale, and in the remaining cases by two judges using an early form of the scale. The two judges differed by more than one step of the scale in only three samples out of ten. Thus the combined opinion of the two judges, though giving only a rough estimate for any single sample, is sufficiently precise for estimating the average quality of the writing of a group of thirty or more pupils, such as a school class.

Section 8. Differences between Systems

It is known that school systems differ greatly one from another in arithmetic (Stone, '07) and much less in spelling (Rice, '97, Cornman, '02). They differ markedly in handwriting if we compare them for its quality, but six of the seven differ hardly any when they are compared for quality of writing done at the same speed.

Thus in the case of the eighth grade the median quality (for the whole grade) of writing done "as well as you can" varies from 11.4 of the Thorndike scale in system A to 14.5 in system F (see Table V); the median quality for the whole grade of writing done "at your usual rate" varies from 10.3 in system A to 14.0 in system G (see Table V).

In average speed of the eighth-grade writings, there is a range from system G with 37 letters a minute to system B with 67

[1] To whom I am greatly indebted for permission to use this material. The ratings by the six judges were also obtained by him.

letters per minute in the careful writing, and, in the writing at one's usual rate, a range from system G with 50 letters per minute to system C with 80. (See Table V.)

TABLE V

SPEED IRRESPECTIVE OF QUALITY AND QUALITY IRRESPECTIVE OF SPEED IN THE CASE OF THE HANDWRITING OF SEVEN SCHOOL SYSTEMS. MEDIAN RESULTS FOR EIGHTH-GRADE PUPILS (UPPER FIGURE) AND SEVENTH-GRADE PUPILS (LOWER FIGURE).

School system	A	B	C	D	E	F	G
Letters per minute in careful writing	61 46	67 48	66 62	51 46	48 43	47 48	37 26
Letters per minute in "natural" writing	79 68	75 63	80 75	60 61	63 56	63 64	50 36
Quality (by T. scale) in careful writing	11.4 11.4	11.8 11.7	11.7 11.4	12.0 11.7	12.3 13.0	14.5 14.0	14.1 13.1
Quality (by T. scale) in natural writing	10.3 10.3	11.2 11.1	11.0 10.6	11.7 11.3	11.6 12.7	13.7 13.0	14.0 13.0

If we compare the eighth-grade pupils of the different schools who wrote at roughly the same rates, they range as shown in Table VI.

TABLE VI

QUALITY OF HANDWRITING AT ROUGHLY THE SAME RATE IN SEVEN SCHOOL SYSTEMS

Median Results for Eighth-Grade Pupils

System	A	B	C	D	E	F	G
At 20-29 words in 4 minutes				14.5	13.0	15.4	14.8
At 30-39 words	11.5	11.3	11.6	12.3	12.3	14.5	14.2
At 40-49 words	11.5	12.0	12.0	11.8	12.3	14.4	15.3
At 50-59 words	11.5	11.6	11.1	11.1	11.6	13.0	11.7
At 60-69 words	10.3	11.8	11.5	11.3	11.6	13.6	
At 70-79 words	10.0	10.8	11.3				

Median Results for Seventh-Grade Pupils

System	A	B	C	D	E	F	G
At 10-19 words					13.3	14.5	13.5
At 20-29 words	12.3	13.3		13.0	13.6	14.2	13.0
At 30-39 words	11.0	11.8		12.3	13.3	14.2	13.0
At 40-49 words	11.0	11.8	11.3	11.7	11.0	13.3	11.8
At 50-59 words	10.3	11.4	11.1	11.0	11.8	13.0	
At 60-69 words	10.0	11.3	10.5	10.0	11.4	11.8	
At 70-79 words	9.8	9.8	9.9				

That these likenesses and differences between systems are not due to chance but are true characteristics of the school systems, is proved by the fact that the measurements come from so many cases as to be subject to only very slight chance or accidental

errors, and also by the slight differences found between buildings or classes within one system. For instance, of three schools of system F, the seventh grades of the two most unlike differ in quality at the same rate by only three tenths of a step, the best one being about 3 per cent better than the worst. Of three schools of system G, the seventh grades of the two most unlike differ in quality at the same rate by only four tenths of a step, or 4 per cent. The average difference of one school from another within the same system is little if any over two tenths of a step or 2 per cent.

The comparison of systems by the quality of the handwriting at the same rates is not, however, perfectly just. For in so doing we are comparing the more rapid writers of some systems (especially G) with the slower writers of other systems (especially A, B, and C). Now in general the more rapid writers of a system would write a little better at the average rate for that system than would the slower writers of the same system. If, for instance, all the seventh-grade pupils of any systems were made to write at different speeds until for each one a record at the rate of 64 letters per second was secured, we should find that the pupils whose undirected rates were higher than 64 would at the 64 rate do better than those whose undirected rates were below 64.

Also, within the 10-19 or 20-29 or 30-39 word group, there are prejudicial differences between the schools. In grade 8 of school G, for example, the 20-29 word group will contain a larger percentage of pupils writing at rate 20-25 than it will in grade 8 of school F. In comparing a slow-writing school with a fast-writing school, by the quality of groups roughly equal in speed, we thus give an unfair advantage to the slower writing school.

To make the comparison just we would have to find equivalents in quality for each degree of superiority in speed, so as to be able to combine the two into a score for general efficiency in handwriting for each individual. Such an investigation of the exact relative values of certain increases in speed and in quality should some time be made. I have not had time or means to make it as yet. As our data are, in the case of the 8th grade pupils, systems A, B, and C can be compared *inter se* for they are substantially alike in speed. So also can systems D, E, and

F. A, B, and C can be compared with D, E, and F with only slight chance of error by taking the "careful" work of A, B, and C and the "natural" work of D, E, and F. But the selection of cases at equal rates slightly favors D, E, and F at the expense of A, B, and C and favors G at the expense of all the other six systems.

Of systems A, B, and C in the eighth-grade writing, A is the worst by about 5 per cent. Of systems D, E, and F, F is the best by about 25 per cent, D and E being alike. F is equal to G in quality and nearly 30 per cent superior in speed. Roughly estimating the equivalence of A's greater speed and F's superior quality, F seems to be about 30 per cent better than A. A, B, C, D, and E are of about equal merit. The data to support these comparisons were given in Tables V and VI.

The records from the seventh-grade pupils give substantially the same result as those from the eighth-grade pupils just stated. A, B, C, D, and E differ little, A being about nine tenths as good as the others. F is about 25 per cent better than B, C, and D. E is a little better than B, C, and D. F shows writing as good as G's at a speed nearly six sevenths greater. These facts are derived from the data of Table VI, an allowance being made for the two constant sources of injustice in such data, and from the data of Table V, G seems, speed and quality both being taken into account, to be little or no better than B, C, D, and E.

Section 9. The Relation of Differences in Results to Differences in Means and Methods of Teaching Handwriting

Not much can be proved by relating these differences to differences in means and methods of teaching handwriting, since the number of school systems studied is so few. F, which is so markedly superior, uses vertical writing of a special system arranged by the supervisor of handwriting, uses writing books, devotes 75 minutes weekly to specific instruction and practice in writing in grades 5, 6, and 7 (of what is done up to grade 5, I have no report), and 30 minutes weekly in grade 8. The teachers in general follow the same system in writing on the blackboard.

The other systems are about alike in general merit in handwriting, A, B, and C gaining speed at a reasonable cost in

quality. A and B teach no fixed system, devote no time to penmanship as such, and permit the teachers to write according to any or no system. C uses a medium slant or intermediate or business system, uses copy books, devotes 50 to 60 minutes weekly to penmanship as such, and has the teachers use the system taught to the pupils. D uses a modified Spencerian with copy books, devotes 50 to 75 minutes weekly in grades 5 and 6, and 75 to 100 in grades 7 and 8, to penmanship as such, and has the teachers follow the system in their own writing. E uses a forward slant, and devotes 100 minutes in grade 6, 60 to 90 in grade 7, and 60 in grade 8, to penmanship as such.

System G uses an intermedial writing, devoting from 60 to 90 minutes weekly to special instruction and practice in grades 7 and 8. The teachers use the same system in their own writing on the blackboard.

What these facts do prove is: First, that at least three systems (C, D, and E) get little or no better results at a time cost of about 75 minutes a week than two systems (A and B) do at zero time-cost; second, that one system (F) at no greater time-cost than C, D, and E gets results about 25 per cent better than they do; and third, that practice for quality may secure it only at the cost of speed. The teachers in A and B are better paid than those in the other cities, so that the success of these schools at no time-cost might not be generally attainable.

Leaving F out of account, the differences of these school systems in the method of teaching handwriting, in the time devoted to it, and in the ideals of the system in respect to it are of inconsiderable influence upon efficiency. One makes its pupils write very well at very slow rates, the others vary a little in quality with small inverse variations in speed. On the whole, in spite of the achievement of system F, efficiency in handwriting seems, like spelling, and unlike arithmetic, to be under present conditions not very much influenced by the management of the schools.

Section 10. Differences between Individuals within the Same School System

We have seen that the school systems, with the exception of F, differ little among themselves in the efficiency of the handwriting which they secure. Individual pupils on the contrary

do differ greatly. Excluding system F, we still find amongst eighth-grade pupils a range from a pupil writing only 55 letters per minute at quality 7.5 up to a pupil writing 79 letters per minute at quality 15.2. Of the 15 eighth-grade pupils writing at the same rate (53 words in 4 minutes) the scores for quality (excluding system F) run from 9.1 to 14.6.

The variation among pupils of the same grade in the rate of writing of the same quality is also large. In the seventh grade of system F, for instance, of the writings of the pupil's usual rate there were 74 samples of from quality 12.7 to 13.3. The number of words written in four minutes ranged from 19 to 87. The 87 may possibly be due to overtime writing. The next highest case was 77, and the next 70. After the two cases at 19

TABLE VII

RELATIVE FREQUENCIES OF DIFFERENT SPEEDS OF WRITING AT NATURAL RATE, THE SAME QUALITY OF WRITING BEING SECURED

Number of words	Frequencies F, 7th grade	G, 7th grade
15		2
17		
19	2	
21		2
3		1
5	2	2
7	1	2
9	4	2
31	2	2
3	4	
5	3	
7	5	
9	3	1
41	3	
3	5	
5	5	
7	2	
9	4	1
51	0	
3	4	
5	5	
7	6	
9	5	
61	3	
3	3	
5	0	
7	1	
9	1	

Also 1 case at 77 and 1 at 87.

there is a gap till 25. From 25 to 70 there is a fairly continuous distribution. Allowing for the time of reading the copy and other disturbing factors, it still appears certain that within the same grade some pupils spend at least three times as long in writing the same amount at the same quality. The facts in detail appear in Table VII.

Individual pupils within the same grade then show a range of difference much greater than that between the fifth grade of the worst system and the eighth grade of the best.

Section 11. The Relation amongst Individuals between Speed and Quality

Rapidity is in and of itself a good sign. If we know nothing about one score or so of pupils save that they are rapid writers and nothing about another score save that they are slow writers, we can prophesy that at the same rate the former group will on the average do writing of a higher quality.

Thus there were 20 seventh-grade pupils in system F who, in writing naturally, varied from 29 to 64 words in four minutes, but who wrote just the same number of words (33) in the test in careful writing. The naturally slower ten showed at the rate of 33 a median quality of 14.5, the more rapid ten a quality of 14.8. This occurred in spite of the fact that for the slower ten writing at a rate of 33 was more like their usual habits. It is to be expected that, if all had been made to write at, say, 73 words in four minutes, the difference would have been greater. Dividing three similar sets each into a slower and a faster half we find for the naturally slower half a median quality of 13.9 and for the faster half a quality of 14.2 when both wrote at an identical rate.

Of course the *same* pupil will not write as well at a rapid as at a moderate rate, and if we mix pupils from a school in which rapidity is gained at the cost of quality with pupils from a school in which quality is gained at the cost of speed the rapid writers will *seem* to be the poor writers. But, in and of itself, rapidity is a sign of ability which if directed toward quality could secure high results in that instead.

Section 12. The Relation of the Quality of Slow Writing to the Quality of the Rapid Writing by the Same individual

Amongst school children there is a close relation between the quality of writing at a natural rate and that at a slower rate. For instance, let us take the 26 children in grades 6 and 7 of one school of system F, who write at a rate of 33-37 inclusive in the first test and at a rate of 52-60 inclusive in the second test and ask whether high rank for quality of writing at a slow rate involves high rank for quality of writing at a speed some 60 per cent greater. It does, the average correlation for three groups like the one described being about .6. Part of this correlation is due to the differences of the individuals in maturity, but this spurious correlation is offset by the attenuation due to the chance variations in the measures related. Of course, the same relation is not thereby proved to hold also between quality of writing at one's natural rate and quality of writing at a rate much *slower* than it. But I believe that it would, and that, although writing at one rate is not identical with writing at a different rate, writing is, over a wide range of rates, so similar a function that training which improves its quality at any one rate may be expected to be of benefit at many others. This does not of course make it any less desirable to practice handwriting at the rate at which one will have to write, but it does make the custom of slow, elaborate writing less pernicious than it might be were the habits at different rates almost or quite independent of one another.

Section 13. Miscellaneous Comments

The Effect of Reduction from a Pupil's Natural Rate upon the Quality of his Handwriting

The gain in quality which a pupil secures by writing more slowly than his natural rate is not great. Sixty-one pupils, whose natural rate was from 52 to 58 words in four minutes, by reducing their speed to 32 to 36 words in four minutes, that is, by writing only two thirds as fast, gained on the average in quality less than one step of the scale. The loss in quality which a pupil suffers by writing more rapidly than his natural rate has not been measured.

The Significance of the Inferiority of Adults' Handwriting

That children in the last two grammar grades can write so much better than adult women-teachers do customarily write is an important fact. Considering it in connection with the fact that above quality 11 there is very little difference in legibility, one is tempted to advocate the heresy that children are taught to write too well. I personally do advocate it. If school boards would furnish, for the use of children electing "writing" as a study in the last two grammar grades, typewriting machines, I should certainly advise the transfer to typewriting of a child in these grades whose writing at 60 letters a minute consistently reaches quality 13. For, the amount of practice required to advance such a pupil to quality 16 at a rate of 75 letters a minute would much more than suffice to advance him to substantially errorless machine writing at that rate. The value now attached to the high qualities of handwriting is, of course, largely fictitious. Employers who can afford such high qualities of writing, buy machines to produce them. For writing cash checks, simple book-entries, labels, and the like, a good plain hand or our quality 12 is entirely adequate. For attaining the higher qualities (15-18) the machine is a more economical tool than the pen, and in my opinion should be provided by those schools which require such qualities. Further, such qualities should, in my opinion, be required of children in the elementary schools, only when they have elected writing as a vocational subject. For the data from the adult women-teachers make it practically certain that the ability to write above quality 14 will not be exercised in life except as a part of a clerical trade. If very, very few teachers find it worth while to maintain qualities above 14, it can hardly be supposed that it will be worth while for mechanics, housekeepers, farmers, and dressmakers to do so.

It seems likely also that handwriting has been, and is, a case of a common practical fallacy, which may be called "learning for learning's sake." When certain facts or acts of skill are teachable, teachers tend to teach them regardless of any intelligible service performed by them other than the doubtful one of "disciplining" the mind or hand or eye. Since, for instance, arithmetical methods of extracting cube root have been learned by

teachers and can be taught to children, we teach them, regardless of the fact that no person in his senses would extract a cube root in that manner. Similarly it is doubtful if any intelligent person would (except to become a teacher of handwriting!) pay the necessary time-cost to acquire the ability to write at 75 letters or over per minute at quality 17 or better. He would, of course, learn to typewrite instead. And if an intelligent person has been artificially induced to get that ability in school, he promptly loses it thereafter.

The Relation between an Individual's Ability in Handwriting and his General Intellectual Ability

I have measured the correlation between scholarship and quality of handwriting in the case of adult women students in Teachers College. These students comprise in the main teachers of from 1 to 15 years of experience who have left their work temporarily for further academic and professional training. The academic marks represent intellect more often than college marks in general do, because what intellect a student has is more likely to be devoted to scholarship in a professional school than in a college. But, on the other hand, achievement in professional courses for teachers is probably aided by experience in teaching. Finally, the standards of marking vary with instructors, and probably somewhat with departments, so that two students, each taking the bulk of her work in some one department, may for equal scholarship receive different average grades.

The marks for quality of handwriting on the other hand are very exact. For each student I have from one to three judgments in each of ten samples taken at random from her writing, — in all a total for each student of nineteen judgments. The marks for handwriting have in fact probable average deviations of the true from the obtained averages of only about 3 per cent.

Gesell[1] has claimed that in children there is a high positive correlation amongst individuals between accuracy of handwriting and intellectual ability. His own data, however, really show a correlation of only about .3, and it seems probable that what

[1] Accuracy in Handwriting, as Related to School Intelligence and Sex. *Am. Jour. of Psy.* Vol. 17, pp. 394-405.

correlation there is is due in large measure to the greater zeal of the brighter pupils to excel in all respects, rather than to any fundamental close correspondence of intellect with motor ability.

The records of our adults add to the evidence that ability in thought and ability in movements are, in adults, only very slightly related. The correlation between scholarship grade and quality of handwriting is zero. For 21 individuals taken at random from the best third of the writers, the median grade in scholarship was about exactly C +. For 22 individuals taken at random from the worst third in writing the median grade in scholarship was a little above B—.

Legibility as a Factor in the Quality of Handwriting

Legibility is a word with many possible meanings. In one of its meanings the legibility of writing may be measured by the distance at which it can be read with a given accuracy and rate. In another of its meanings it can be measured by the rate at which a sample can be read at a given distance and with a given accuracy. Still finer discrimination might perhaps be made between samples of handwriting equally legible by these two tests, by a further test of the degree of fatigue or of discomfort resulting from reading them. Even the subjective measurement of legibility by the combined opinions of competent judges is useful.

Legibility may be tested in any one of these ways in the case of single letters or groups of letters cut from a sample, of words alone or in a random order, and in the case of words in a sensible context where what is before and after the word helps one to read it. Also, legibility for a tyro at reading and for a trained reader may well be different things, requiring separate investigation.

I have made rough measurements of the legibility of the different qualities of the scales for children's and adults' writing, but the results are too meagre for publication at this time. They show, as might be foreseen, that the higher grades of quality differ little in the legibility of the words in a context for a trained reader, but correlate fairly closely with the legibility of the letters singly.

Section 14. A Scale of Children's Writing, Based on Equally Often Noted Differences in Quality

In section 2 the fact was noted that a series of samples, *a, b, c, d,* etc., could be found such that when each was compared with the others *a* was judged better than *b* as often as *b* was judged better than *c,* and so on. It was pointed out that, if the per cents of judgments of *a-b, b-c,* etc., are under 100, over 50, and equal, the samples *a, b, c,* etc., may be regarded as differing by equal steps, equal differences meaning *equally often noticed* differences. A scale constructed on this principle differs from one constructed on the principle that differences judged equal by the combined opinion of competent judges are equal, chiefly in respect to the lower qualities of handwriting. The difference between lower qualities that is *noticed as often* as a given difference between higher qualities is regarded by competent judges as a *smaller* difference than the latter.

The reasons for this are interesting and the general problem of the choice between "judged equal" and "noticed equally often" as the basis for equality of units in psychological, educational and sociological scales is of very great importance. But much more elaborate investigations than I have been able to make are necessary for any adequate discussion of the problem, even in the case of handwriting alone.

I therefore simply present the scale as derived by this second method for the use of those who prefer it. The facts are that:

Sample 125 is, in direct comparison with sample 132, judged to be better by 81 per cent of 42 judges.

132	is	similarly	judged	better	than	58	by 81%	of 42 judges
58	"	"	"	"		10	by 79%	"
10	"	"	"	"		1	by 79%	"
1	"	"	"	"		55	by 79%	"
55	"	"	"	"		52	by 77%	91
52	"	"	"	"		44	by 83.5%	"
44	"	"	"	"		34	by 79%	"
34	"	"	"	"		51	by 78%	"
51	"	"	"	"		22	by 80%	"
22	"	"	"	"		126	by 74%	42

Samples 126, 22, 51, 34, 44, 52, and so on, thus have approximately the values x, $x + k$, $x + 2k$, $x + 3k$, $x + 4k$, respectively, where $x =$ the difference in quality between sample 126 and zero

quality, and k = such a difference in quality as is noted as different by competent judges eight times out of ten. K is, on the average, equal to 1 of steps 7 to 17 of scale A, but the value of x remains undetermined. If a series was found ranging from a sample arbitrarily taken as of zero merit (say, sample 140) to sample 126, such that the difference between any two successive samples was noted eight times out of ten, the number of steps in this series would give the value of x in terms of k. X is probably between $8k$ and $9k$. Since, however, the calculation of the distance of x from o by other methods worthy of consideration makes it as low as $6k$ or even $5k$, I shall use $7k$ as the distance of sample 126 from o. Anyone preferring to adhere rigidly to "noticed equally often" as the sole test of equality of difference may use 19.5, 18.5, 17.5, 16.5, etc., in place of the 18, 17, 16, 15, etc., used in the scale.

HANDWRITING

By EDWARD L. THORNDIKE

Professor of Educational Psychology, Teachers College
Columbia University

PUBLISHED BY
Teachers College, Columbia University
NEW YORK CITY
1912

LIBRARY OF CONGRESS
0 021 775 780 0